Pathways, Plasticity, and Weight Decay in the Decision Tree View of Neural Networks

1. Introduction

ReLU neural networks can be viewed as **implicit decision trees**, where each input follows a path determined by activation patterns (which neurons are on or off). Each path corresponds to a specific **synthetic linear mapping** built from shared weights.

During training, these pathways are not fixed. Instead, they are **formed, adjusted, and sometimes repurposed** as the network learns from data. This report explains how that happens, and how factors like weight decay influence the process.

---

2. What is a "Pathway"?

A pathway is a sequence of activation decisions across layers:

- At each layer, ReLU units turn on or off
- This defines a route through the network
- That route corresponds to a specific linear transformation

So instead of a single function, the network behaves like:

> A large collection of possible linear functions, with one selected per input.

---

3. How Pathways Are Reinforced During Training

When training data repeatedly follows similar pathways:

- The same neurons activate
- The same sequence of transformations is used
- Gradients update the shared weights along that route

Over time:

- The mapping associated with that pathway improves

- The pathway becomes **stable for that type of input**

However, unlike a real decision tree:

> Pathways are not independent—they all share the same weights.

This means:

- Updating one pathway also affects others
- Pathways are **coupled and continuously changing**

---

4. Why Pathways Remain Plastic

Even after training begins to stabilize, pathways remain flexible (“plastic”) for several reasons:

Shared Weights

All pathways reuse the same parameters. Changing weights for one pathway slightly alters many others.

Movable Boundaries

Small changes in weights can flip ReLU activations:

- Inputs can switch from one path to another
- The structure of the “tree” itself shifts

Gradient-Based Learning

Training tends to:

- adjust existing pathways gradually
- reuse them rather than create entirely new ones

So instead of fixed branches, we have:

> A continuously deforming decision tree.

---

5. The Role of Weight Decay

Weight decay (L2 regularization) adds a penalty for large weights. It does not directly remove pathways, but it influences how they evolve.

5.1 Early Training

At initialization:

- Pathways are effectively random
- Training quickly aligns some pathways with real data

Weight decay helps by:

- reducing unnecessary large weights
- encouraging simpler, smoother mappings

However, most “cleanup” of random pathways comes from:

- data-driven learning (SGD), not decay alone

5.2 Without Weight Decay

If weight decay is weak or absent:

- Weights can grow large
- Early structures may persist
- The network can form sharp, highly specific pathways

This can lead to:

- strong memorization
- less flexibility for new data

5.3 With Weight Decay

With stronger decay:

- Weights are kept smaller
- Pathways become smoother and less extreme
- Unused or weak pathways gradually fade

In effect:

> Weight decay encourages the network to simplify and share structure across pathways.

---

6. Fine-Tuning: Preserving vs Replacing Pathways

When training on new data (fine-tuning), there is a key trade-off.

Low Weight Decay

- Existing pathways are mostly preserved
- New data adjusts them slightly
- Old knowledge is retained

→ Good for transfer learning and stability

High Weight Decay

- Old weights shrink
- Existing pathways weaken or disappear
- New pathways can form more easily

→ Useful when adapting to very different tasks

---

7. Pathways and Information Flow

From the decision tree perspective:

- A classical tree always has access to the original input
- A neural network only has access to what remains in its internal representation

So:

> Pathways can only be formed and reinforced based on information that survives through the layers.

If information is lost early:

- Different inputs may become indistinguishable
- Fewer distinct pathways can exist
- The effective tree becomes smaller

Good information flow allows:

- richer pathway diversity
- better conditional decisions
- stronger compositional computation

---

8. A More Accurate Mental Model

Instead of thinking:

> "Training locks in certain paths"

It is better to think:

> "Training sculpts a landscape of possible paths, where frequently used ones become stable and rarely used ones fade or are reused."

---

9. Conclusion

In the decision tree view of neural networks:

- Pathways correspond to activation patterns and associated linear mappings
- Training reinforces pathways by repeatedly updating shared weights
- Pathways remain plastic due to shared parameters and movable decision boundaries
- Weight decay controls how much old structure is preserved versus simplified
- Information flow determines how many distinct pathways can exist and be learned

Overall:

> A neural network does not learn a fixed tree, but a **flexible, evolving system of pathways** shaped by data, optimization, and regularization.